Gesture Segmentation Using Spiking Neural Networks Inspired by the Visual System

L.P. Huang, Q.X. Wu, G.R. Zhang, X. Wang

Key Laboratory of OptoElectronic Science and Technology for Medicine of Ministry of Education
College of Photonic and Electronic Engineering, Fujian Normal University
Fuzhou, China

Abstract—The human visual system demonstrates powerful image processing functionalities. Inspired by the behaviour of the human visual system, a method of gesture segmentation with the complex background is investigated based on fusion of outputs from two kinds of spiking neural networks. The structures and the properties of the two networks are detailed in this paper. One of two networks performs the movement region of gesture. Meanwhile, another spiking neural network performs identification of skin area according to the biologically colour processing mechanism. Finally, two outputs from two neural networks are integrated to segment the meaningful region of gestures from the video image sequence. Simulation results show that the proposed algorithm works efficiently and can perform gesture segmentation with the satisfying accuracy for dynamic visual image sequence under complex background. It is promising to apply this approach to video processing domain and robotic visual systems.

Keywords-gesture segmentation; spiking neural network; visual system; motion detection; skin detection

I. Introduction

The research of segmenting a gesture accurately from a complex background in a sequence of images is very important for understanding of human gestures. It is the most crucial step in every gesture recognition system [1]. Accordingly, many scholars have already proposed a lot of methods which are based on the computer vision to improve the segmentation quality of gestures. It can be seen that all the mentioned techniques in [2-4] utilize several complex algorithms and multiple sensors. In this paper we propose a new method based on two kinds of spiking neural networks inspired by the visual system to segment gestures from images in a video.

The human visual system demonstrates powerful image processing functionalities to extract key information of different objects from the complicated environment. For example, a lot of researchers have studied simulations of the human visual system, which is called artificial vision system [5]. Based on the receptive fields inspired by biological system, different spiking neural network models were used to edge detection, color image segmentation [6, 7] and feature extraction [8]. The visual system, in which spiking neurons are basic units for information processing, has powerful intelligent functions to process visual images. It can be boldly predicted that most image processing functions can be implemented using different spiking neuron-based networks. Neuroscientists have found that there are various receptive fields from simple cells in the striate cortex to those of the retina and lateral geniculate nucleus [9], and the axonal delay causes a phase shift for a spike train [10-14]. Jeffress [15-16] applied the time difference principle of axonal delay to account for sound localisation [11, 13]. Biological evidence also shows that the retina contains two forms of photosensitive neurons, i.e. rods and cones [17]. Rod neurons are highly sensitive to light and respond in dim light and dark conditions. Three types of cones have primary sensitivity of red, green, and blue lights. In this paper, a spiking neural network model is proposed to detect motion gesture in an image sequence based on axonal delay [10-14]. Meanwhile another neural network is explored to identify skin area of gesture in an image using color processing mechanism inspired by the visual system, in which grey scale, colour RGB information and HSV information are integrated. The meaningful gesture can be extracted by the fusion of the movement and color information in video image sequences. The output is encoded by the fire rate of the neuron in the each network. Adjusting the firing threshold value of spike neuron, the noise and some unimportant information can be ignored. Therefore, accuracy of segmentation of gesture can be improved.

The main contribution of this paper is to presents a new type of neural network for skin detection, which is different from the neural network for colour segmentation [19]. In addition, fusion of two outputs from two neural networks inspired by the biological visual system is used to segment gesture from a video image sequence of a video, which may help us to understand how the visual system extracts the important information from the image.

The remainder of this paper is organized as follows. Section 2 focuses on the algorithm for motion gesture detection based on the spiking neural network model. The proposed method for skin detection inspired by the visual system is described in Section 3. In Section 4, fusion of two outputs from the two neural networks is introduced. The experimental results are presented in Section 5 and a conclusion is given in Section 6.

II. Spiking Neural Network Model for Motion Gesture Detection

Inspired by the axonal delay mechanism, a spiking neural network model is proposed to simulate the function of difference method to extract the motion region of gesture from dynamic visual image sequence. The architecture of neural network is shown as in Figure Ⅰ. Its structure is shown in Figure. Ⅰ

FIGURE I. SPIKING NEURAL NETWORK MODEL FOR MOTION GESTURE DETECTION.

The first layer represents photonic receptors for an image from visual system. Each pixel corresponds to a receptor. The intermediate layer is composed of two neuron arrays of same size. The third layer represents the output layer. Each neuron in the output layer integrates corresponding outputs from intermediate neurons. As shown in Fig.1, the neuron Nr(x, y) is connected to N1(x, y) through excitatory synapse without delay and through inhibitory synapse with an axonal delay Δt , while N2(x, y) connects to the neuron Nr(x, y) on the contrary. If pixel Nr(x, y) is on the moving target, the gray scale of pixel in the image will change through axon delay and the current from receptor Nr(x, y) will be unstable. So Neuron N1 or Neuron N2 will generate spikes. Finally, Neuron N(x', y') in the output neuron layer will fire. Therefore, the moving object corresponds to high firing-rate neurons in the output layer. The object can be obtained by binding these highly active neurons in the output layer. More details can be found in [18].

The block diagram of the proposed system for motion gesture detection is shown in Figure. Ⅱ. The system takes an gesture image from the image sequence each time step. The gesture image is transferred to a grey scale image. The grey image presents to the spiking neural network (SNN) for motion detection. The moving gestures can be detected by the SNN. The edges of firing neuron groups are used to determine the boundaries of the moving gestures. Using the boundaries of the gestures, a filter is generated to take out of moving gestures from background. Therefore, the moving gesture in the grey image is transferred to the output image.

FIGURE II.BLOCK DIAGRAM OF THE PROPOSED SYSTEM.

III.   NEURAL NETWORK MODEL FOR SKIN DETECTION

Inspired by the roles of rods and cones in retina of the human visual system, the neural network has proposed to simulate the function of integration of colours, grey information in a colour image. The H component in HSV can only express the chromatic information with little influence. Based on the neural network model [19], the H component have been added to perform perfect skin detection. This will enhance capability of the network to distinguish objects using property of H component. The architecture of neural network is shown as in FigureⅢ. The output value of neuron O(x, y) is determined by two receptive fields centred at P(x, y) in the input image. RCx,y is a 3*3 receptive field simulates three types of cones corresponding to color components R, G, and B. The 5*5 RNx,y receptive field simulates rod receptive field corresponding to gray scale. Neuron H in the input layer is designed to simulate hue neuron. The five neurons are fully connected to a hidden neuron layer. All neurons in the hidden neuron layer are connected to neuron O(x,y).
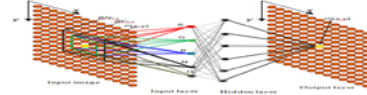

FIGURE III. ARCHITECTURE OF NEURAL NETWORK MODEL.

Let r(x',y'), g(x',y'), b(x',y') and n(x',y') represent values of colours R, G, B, and gray-scales of pixel P(x',y') in the input image respectively. According to the receptive fields in the network architecture, we have

$$R_{x,y} = \frac{1}{A} \sum_{(x',y') \in RC_{x,y}} r(x',y') \quad , \quad (1)$$

$$G_{x,y} = \frac{1}{A} \sum_{(x',y') \in RC_{x,y}} g(x',y') \quad , \quad (2)$$

$$B_{x,y} = \frac{1}{A} \sum_{(x',y') \in RC_{x,y}} b(x',y') \quad , \quad (3)$$

$$N_{x,y} = \frac{1}{B} \sum_{(x',y') \in RC_{x,y}} n(x',y') \quad , \quad (4)$$

where A is area of the receptive field RCx,y for cone neurons, B is area of the receptive field RNx,y for rod neurons, and gray scale n(x',y') = 0.2989 r(x',y') + 0.587 g(x',y') + 0.114 b(x',y'). In the experiments, dimension of receptive field RCx,y is set to 3×3 and dimension of receptive field RNx,y is 5×5. Therefore, A=9 and B=25. These receptive fields can be adjusted according to different images. Segmentation based on only the primary colors might lead to false results due to the fact that the background may also have a similar color combination. Hence color conversion is extremely essential. Neuron Hx,y is used to detect the hue around point (x,y). Suppose that the response is simulated by following expression.

$$H_{x,y} = \frac{1}{A} \sum_{(x',y') \in RC_{x,y}} h(x',y') \quad (5)$$

Where h(x',y')is calculated by following expressions[20].

$$h_1 = \cos^{-1} \frac{0.5 \times [(r(x',y') - g(x',y')) + (r(x',y') - b(x',y'))]}{\sqrt{(r(x',y') - g(x',y'))^2 + (r(x',y') - b(x',y')) \times (g(x',y') - b(x',y'))}} \quad (6)$$

$$h(x',y') = \begin{cases} h_1, b(x',y') \le g(x',y') \\ 360^0 - h_1, \ b(x',y') > g(x',y') \end{cases} \quad (7)$$

Let $(R_{x,y}, G_{x,y}, B_{x,y}, N_{x,y}, H_{x,y})$ represent input vector of the network and O(x,y) represent output value of the output neuron. The output of the network can be represented by the following expression.

$$O(x,y) = NET(R_{x,y}, G_{x,y}, B_{x,y}, N_{x,y}, H_{x,y}) \quad (8)$$

O(x,y) can be defined as a classification numbers. For example, in this paper it is defined as three numbers {0, 1, 2}. "0" represents background. "1" represents skin pixels. "2" represents non-skin pixels. The simulation results are in the next section.

IV. Fusion of Two Outputs from Two Neural Network

After motion detection is carried out by the spiking neural network model in figure.1, O1 represents the output of the movement gesture. Simultaneously, on the other side skin color segmentation is done to segment out the skin region by the model in figure.3, O2 is the output of the gesture skin. In the next step logical AND operation is performed between outputs O1 and O2 to get the final hand segmented output. The fusion can be express as:

$$hand = \begin{cases} 1, O1 = 1 \ \& \ O2 = 1 \\ 0, \ other \end{cases} \quad (9)$$

V. Experimental Results

A. The Results of Motion Gesture Detection

Using the Spike Neural Network architecture mentioned above, the movement region of gesture can be detected. Figure. IV(a) is the collected original gesture image, which is 320x240 in size. Figure. IV(b) is the binary motion target image of gesture.
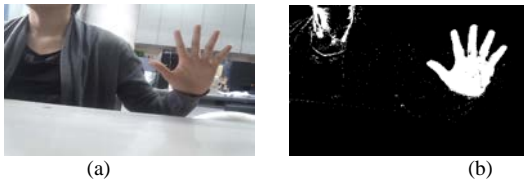

| (a) | (b) |

FIGURE IV. (A) THE COLLECTED ORIGINAL GESTURE IMAGE (B) THE BINARY MOTION TARGET IMAGES OF GESTURE.

B. The Results of Skin Detection

Meanwhile through the above neural network, the skin color region of gesture is obtained. Figure. V (a) is the collected original gesture image. Figure. V (b) is the binary skin color image of gesture. Figure. V (c) is the separated skin areas.
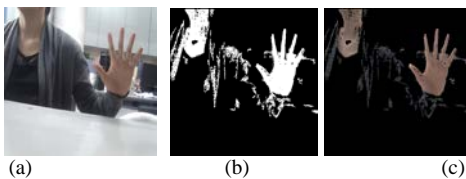

| (a) | (b) | (c) |

FIGURE V. (A) THE COLLECTED ORIGINAL GESTURE IMAGE(B) THE BINARY SKIN COLOR IMAGE OF GESTURE (C) THE SEPARATED SKIN AREAS.

C. The Results of Fusion

Then the gesture region can be gained by the AND operation on the movement region and the skin color region. The following extracted images of a sequence will show the results of gestures segmentation in complex background. Figure.VI is the ultimate number gesture segmentation image after morphology processing.

FIGURE VI. THE GESTURE SEGMENTATION IMAGES.

VI. Conclusions

This thesis has been focused on solving the challenging issues that automatically segments from continuous video streams. In particular, a novel gesture segmentation approach has been proposed, which has been shown to be very effective. The results of spiking neural network for motion detection show SNN can quickly obtain the main features and remove non-significant information and noise through fire threshold adjustment. There are few restrictions for the gesture feature extraction in the face of complex backgrounds. This paper also presents a neural network for skin detection from a colour image. The colour processing mechanism inspired by the visual system is used in the network. The advantage of the proposed algorithms is that the information of colour, gray has been integrated to identify colour objects in an image. In this method, there is no need to think of the distribution region of skin colors in a color space. With the abundant training samples, the neural network is capable to characterize the distribution region of skin colors accurately in the color space and identify the skin regions efficiently in the color image with complex backgrounds. Based on the fusion of two network, a hand gesture can be extracted in an image from video image sequence. It is promising to apply this method to improve the gesture recognition further.

Reference

[1] Choudhury, A.; Talukdar, A.K.; Sarma, K.K., "A novel hand segmentation method for multiple-hand gesture recognition system under complex background, " on Signal Processing and Integrated Networks, pp.136-140, 20-21 Feb. 2014.

[2] A. Chaki, P. Jain, and R. K. Gupta, "Information Measure Ratio Based Real Time Approach for Hand Region Segmentation with a Focus on Gesture Recognition," 2nd International Conference on Intelligent Systems, Modelling and Simulation (ISMS), pp. 172 – 176, 2011.

[3] N. Dhruva, Sudhir Rupanagudi and H. N. Neelkant Kashyap, "Novel Algorithm for Image Processing based Hand Gesture Recognition and its Application in Security," Third International Conference on Advances in Computing, Communication, and Control (ICAC3), Communications in Computer and Information Science, Volume 361, pp. 537 – 547, January 18-19, 2013.

[4] Shu Mo. Shihai Cheng and Xiaofen Xing Sch. of Electron. & Inf. Eng., South China Univ. of Technol., Guangzhou, China, "Hand gesture segmentation based on improved kalman filter and TSL skin color model," International C.

[5] A. Bharath and M. Petrou, Next generation artificial vision system, reverse engineering the human visual system, Artech House, 2008.

[6] Q.X. Wu, T.M. McGinnity, L. P. Maguire, A. Belatreche, and B.Glackin, "Edge Detection Based on Spiking Neural Network Model", Springer-Verlag Berlin Heidelberg, LNAI 4682 pp.26–34, 2007.

[7] Q. Wu, X. Huang, J. Cai, Y. Wu, M. Lin, Segmentation of Leukocytes in Blood Smeare Images Using Color Processing Mechanism Inspired by The Visual System, in: Biomedical Engineering and Informatics, 2009. BMEI'09. 2nd International Conference, 1-4, IEEE, 2009.

[8] Q.X. Wu, T.M. McGinnity, L. P. Maguire, B.Glackin, and A. Belatreche, "Information Processing Functionality of Spiking Neurons for Image Feature Extraction", Seventh International Workshop on Information Processing in Cells and Tissue 2007, 29th - 31st August 2007, Jesus College, Oxford, UK.

[9] Kandel, E. R., Shwartz, J. H.: Principles of neural science. Edward Amold (Publishers) Ltd., (1981).

[10] Lin, J.W., Faber, D.S.: Modulation of synaptic delay during synaptic plasticity. Trends Neurosci. 25(9) 449--55(2002).

[11] Pena, J. L., Kazuo S. VSaberi, F,K., Konishi, M.: Cochlear and Neural Delays for Coincidence Detection in Owls, The Journal of Neuroscience, 21(23), 9455--9459(2001).

[12] Senn, W., Schneider, M., Ruf, B.: Activity-Dependent Development of Axonal and Dendritic Delays, or, Why Synaptic Transmission Should Be Unreliable, Neural Computation. 14, 583--619(2002).

[13] Carr, C. E., Konishi, M.: Axonal Delay Lines for Time Measurement in the Owl's Brainstem, Proceedings of the National Academy of Sciences of the United States of America, 85(21) 8311--8315(1988) .

[14] Crook, S.M., Ermentrout, G.B., Vanier M.C. , Bower, J.M.: The Role of Axonal Delay in the Synchronization of Networks of Coupled Cortical Oscillators, Journal of Computational Neuroscience, Publisher Springer Netherlands, 4(2), 1573--6873(1997).

[15] Jeffress, L. A.: A place theory of sound localization, J. Comp. Physiol. Psychol., 41, 35--39(1948).

[16] Jeffress, L. A.: Binaural Phase Difference and Pitch Variation, Am. J. Psychol., 61, 468--486. (1948)

[17] E. R. Kandel, and H. J. Shwartz, Principles of neural science. Edward Amold (Publishers) Ltd., (1981).

[18] WU Q X, MCGINNITY T M, MAGUIRE L P, et al, Motion detection using spiking neural network model [M].[S. l.]: Springer-Verlag, 2008.

[19] Q.X. Wu, Xi Huang, Jianyong Cai, Yi Wu, Meiyan Lin, "Segmentation of Leukocytes in Blood Smeared Images Using Color Processing Mechanism Inspired by The Visual System", Proceedings of International Conference on BioMedical Engineering and Informatics (BMEI'09), IEEE, pp.368-372,(2009).

[20] C.H. Cui, C.M. Zhu, "Real-time Human Face Detection and Tracking Based on HSV Model Space of Skin Color", Journal of Fuzhou University (Natural Science Edition), vol.34, no.6, 2006.